

\*\*\*\*\*  
 W O R D S (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPearch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Sat May 13 08:34:56 2000; Maspar time 21.20 Seconds  
 659.132 Million cell updates/sec  
 Tabular output not generated.

Title: >US-09-331-631-8  
 Description: (1-590) from US09331631.pep (1 of 4)  
 Perfect Score: 4332  
 Sequence: 1 MVRNKSACVLLFLSLFLSFC.....FNSNPQESFYVRQRRASE 590

Scoring table:  
 PAM 150  
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-geneseq35  
 1:geneseqp

Statistics: Mean 36.568; Variance 172.610; scale 0.212

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	4332	100.0	590	1	Gossypium hirsutum ant	0.00e+00
2	2101	48.5	566	1	Sequence encoded by 67	2.30e-174
3	1910	44.1	525	1	Theobroma cacao anti	4.54e-157
4	1246	28.8	666	1	Macadamia integrifolia	3.18e-97
5	1244	28.7	666	1	Macadamia integrifolia	4.80e-97
6	1236	28.5	666	1	Macadamia integrifolia	2.50e-96
7	966	22.3	605	1	Glycine max antimicrob	3.32e-72
8	955	22.0	593	1	zea mays antimicrob	3.17e-71
9	834	19.3	614	1	peanut allergen Ara hi	1.73e-60
10	834	19.3	614	1	Arachis hypogaea anti	1.73e-60
11	784	18.1	626	1	Peanut allergen Ara hi	4.58e-56
12	761	17.6	409	1	G. max truncated SBP2	4.88e-54
13	764	17.5	489	1	G. max SBP2 protein.	2.65e-54
14	758	17.0	637	1	Hordeum vulgare anti	8.97e-54
15	737	17.0	524	1	G. max SBP1 protein.	6.35e-52
16	716	16.5	444	1	G. max truncated SBP1	4.48e-50
17	190	4.4	395	1	Mouse SRY-related prot	1.65e-05
18	147	3.4	1382	1	Human metastasis-assoc	2.79e-02
19	143	3.3	351	1	Sequence of Histidine-	5.43e-02
20	142	3.3	515	1	Oat globin A2B.	6.43e-02
21	138	3.2	186	1	Tyranosoma cruzi anti	1.25e-01
22	136	3.1	86	1	GST-HD fusion protein	1.73e-01
23	136	3.1	86	1	GST-HD fusion protein	1.73e-01

RESULT ID	Score	Query Match	Best Local Similarity	Conservative	Mismatches	Indels	Gaps
1	4332	100.0%	100.0%	0	0	0	0
2	2101	48.5	100.0%	0	0	0	0
3	1910	44.1	100.0%	0	0	0	0
4	1246	28.8	100.0%	0	0	0	0
5	1244	28.7	100.0%	0	0	0	0
6	1236	28.5	100.0%	0	0	0	0
7	966	22.3	100.0%	0	0	0	0
8	955	22.0	100.0%	0	0	0	0
9	834	19.3	100.0%	0	0	0	0
10	834	19.3	100.0%	0	0	0	0
11	784	18.1	100.0%	0	0	0	0
12	761	17.6	100.0%	0	0	0	0
13	764	17.5	100.0%	0	0	0	0
14	758	17.0	100.0%	0	0	0	0
15	737	17.0	100.0%	0	0	0	0
16	716	16.5	100.0%	0	0	0	0
17	190	4.4	100.0%	0	0	0	0
18	147	3.4	100.0%	0	0	0	0
19	143	3.3	100.0%	0	0	0	0
20	142	3.3	100.0%	0	0	0	0
21	138	3.2	100.0%	0	0	0	0
22	136	3.1	100.0%	0	0	0	0
23	136	3.1	100.0%	0	0	0	0

Query 301 NNGQEEFFPPASQRPQSYLRAFSRILEPFAFTRSEQDELFGGQSRROGQGMFRK 360  
Db 361 ASQEQFRALSQEQATSPREKSGERFAFNLSQTPRYSNONGREFEACPPPEFROLRDNVTV 420  
Qy 361 ASQEQFRALSQEQATSPREKSGERFAFNLSQTPRYSNONGREFEACPPPEFROLRDNVTV 420  
Db 421 SALQLOGSGIFVPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOE 480  
Qy 421 SALQLOGSGIFVPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOE 480  
Db 481 EERRSQYKIRSRISRLSGDIFVVPANFPVTFVASQONLRMTGFGLYNONINPDHQRITF 540  
Qy 481 EERRSQYKIRSRISRLSGDIFVVPANFPVTFVASQONLRMTGFGLYNONINPDHQRITF 540  
Db 541 VAGKINHVMQDMSQAKELAFGVSSRLVDEIFNSNPQESYFVSRRORASE 590  
Qy 541 VAGKINHVMQDMSQAKELAFGVSSRLVDEIFNSNPQESYFVSRRORASE 590

RESULT 2  
ID R20181 standard; Protein: 566 AA.  
AC R20181:  
DT 16-APR-1992 (first entry)  
DE Sequence encoded by 67 kd T. cacao protein cDNA.  
KW Cocoa; flavour: vicillin; seed storage protein.  
OS Theobroma cacao.  
PN WO9119801-A.  
PD 26-DEC-1991.  
PE 07-JUN-1991; G00914.  
PR 11-JUN-1990; GB-013016.  
PI (MRC) MARS UK LTD.  
PA Spencer ME, Hodge R, Deakin EA, Ashton S;  
DR WPI: 92-024418/03.  
DR N-PSDB: 020377.  
PT Recombinant cocoa proteins - are responsible for flavour in cocoa  
PT beans and produced in large quantities using yeast and bacterial  
PT expression vectors  
PS Claim 4; Fig 2; 59pp; English.  
CC The inventors claim a 67 kd and 31 kd T. cacao protein, and  
CC fragments, and encoding DNAs. The 47 kd and 31 kd proteins are  
CC derived from the 67 kd precursor. T. cacao protein cDNA was  
CC detected in a cDNA library prepared from immature cocoa beans RNA  
CC using a probe based on the AA sequence of a CNBr peptide common to  
CC the 47 kd and 31 kd polypeptides. Homology searches revealed close  
CC homologies between the 67 kd polypeptide and the vicillins, which are  
CC seed storage proteins.  
SQ Sequence 566 AA:

Query Match 48.5%; Score 2101; DB 1; Length 566;  
Best Local Similarity 57.8%; Pred. No. 2,30e-174;  
Matches 301; Conservative 104; Mismatches 97; Indels 19; Gaps 13;

Db 37 DPQRYEQOCRCSESEATEERQEOCEQRCERKEQOQOEELQROYQOCQCGQOEQ 96  
Qy 82 DPQRYEQOCRCSESEATEERQEOCEQRCERKEQOQOEELQROYQOCQCGQOEQ 96  
Db 97 QGOREQOOCRCQCEQOEQOE-EGEHENYHNKKNRSEEEGQORNNPYFPRRSFOTR 155  
Qy 136 QRPKKQOCVRECKREKYOENPMRGEREE-FAEEEE-TEEGEOQSHNPFPH-RRSFOSR 192  
Db 156 FRDEGNGKIORFAENSPPLKINDYRLAMEANPNFILPHHCDAEATVINGKGTI 215  
Qy 193 FREDGNGKIORFAENSPPLKINDYRLAMEANPNFILPHHCDAEATVINGKGTI 215  
Db 216 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Qy 252 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Db 276 NNPESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Qy 312 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 312  
Db 313 SORPQSYLRAFASRILEPFAFTRSEQDELFGGQSRROGQ-GMFRKASQEOIRALSO 371  
Qy 371 SORPQSYLRAFASRILEPFAFTRSEQDELFGGQSRROGQ-GMFRKASQEOIRALSO 371

Db 336 QATSPRHGGERLAINLLSQSPVYSNONGREFEACPEDEFSQFQNNDAVAFKLNQGAIF 395  
Qy 372 EATSPREKSGERFAFNLSQTPRYSNONGREFEACPPPEFROLRDNVTVSALQLOGSGIF 431  
Db 396 VPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOEERRSGQYRKI 491  
Qy 432 VPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOEERRSGQYRKI 491  
Db 456 KAPLSPGCVFVAPAGHATFFPASKDQPLNVAFAFGIQAON-----NORIFLAGKKNLVROM 510  
Qy 492 KAPLSPGCVFVAPAGHATFFPASKDQPLNVAFAFGIQAON-----NORIFLAGKKNLVROM 510  
Db 511 DSEAKELSGVPSKLVNDIEN-NDESYFMSFSQORQRDE 550  
Qy 552 DSEAKELSGVPSKLVNDIEN-NDESYFMSFSQORQRDE 550

RESULT 3  
ID W62831 standard; Protein: 525 AA.  
AC W62831:  
DT 27-OCT-1998 (first entry)  
DE Theobroma cacao antimicrobial protein.  
KW antimicrobial protein; infestation; control.  
OS Theobroma cacao.  
PN WO9827805-A1.  
PD 02-JUL-1998.  
PE 22-DEC-1997; AU0874.  
PR 20-DEC-1996; AU-004275.  
PA (RETR)-COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Bower NI, Goullier KC, Green JL, Manners JM, Marcus JP;  
DR WPI: 98-377279/32.  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
PT useful for controlling microbial infestations of plants or mammals  
PS Claim 1; Page 47-49; 96pp; English.  
CC The sequence is that of an antimicrobial protein which can  
CC be used to control microbial infestations in plants and mammalian  
CC animals.  
SQ Sequence 525 AA;

Query Match 44.1%; Score 1910; DB 1; Length 525;  
Best Local Similarity 56.9%; Pred. No. 4.54e-157;  
Matches 261; Conservative 95; Mismatches 92; Indels 11; Gaps 10;

Db 37 DPQRYEQOCRCSESEATEERQEOCEQRCERKEQOQOEELQROYQOCQCGQOEQ 96  
Qy 82 DPQRYEQOCRCSESEATEERQEOCEQRCERKEQOQOEELQROYQOCQCGQOEQ 96  
Db 97 QGOREQOOCRCQCEQOEQOE-EGEHENYHNKKNRSEEEGQORNNPYFPRRSFOTR 155  
Qy 136 QRPKKQOCVRECKREKYOENPMRGEREE-FAEEEE-TEEGEOQSHNPFPH-RRSFOSR 192  
Db 156 FRDEGNGKIORFAENSPPLKINDYRLAMEANPNFILPHHCDAEATVINGKGTI 215  
Qy 193 FREDGNGKIORFAENSPPLKINDYRLAMEANPNFILPHHCDAEATVINGKGTI 215  
Db 216 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Qy 252 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Db 276 NNPESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 275  
Qy 312 TTYTHENKESYVVGQTVSVVAGSTVYVVSQDNOEKLTIAVLALPVNSPKYELFFPAG 312  
Db 313 SORPQSYLRAFASRILEPFAFTRSEQDELFGGQSRROGQ-GMFRKASQEOIRALSO 371  
Qy 371 SORPQSYLRAFASRILEPFAFTRSEQDELFGGQSRROGQ-GMFRKASQEOIRALSO 371  
Db 396 VPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOEERRSGQYRKI 491  
Qy 432 VPHYNSKATFVLVTEGNGYAEWSPHLPROSSYEEDEEEDDEEQOEERRSGQYRKI 491  
Db 456 KAPLSPGCVFVAPAGHATFFPASKDQPLNVAFAFGIQAON 494  
Qy 494 KAPLSPGCVFVAPAGHATFFPASKDQPLNVAFAFGIQAON 494

QY 492 RSRUSRGDIEVPANPVTVAASQNLMTGGLYON 530

## RESULT 4

ID W62828 standard; Protein; 666 AA.  
AC W62828; 27-OCT-1998 (first entry)  
DE Macadamia integrifolia antimicrobial protein.  
KW antimicrobial protein; infestation; control.  
OS Macadamia integrifolia.  
FH key Location/Qualifiers  
FT Peptide 1..28  
FT /note= "signal peptide"  
FT Protein 29..666  
FT /note= "mature protein"  
PN W09827805-A1.  
PD 02-JUL-1998.  
PF 22-DEC-1997; AU0874.  
PR 20-DEC-1996; AU-004275.  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
DR WPI: 98-37279/32.  
DR N-PSDB: V42310.  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
useful for controlling microbial infestations of plants or mammals  
PS Claim 1; Page 34-36; 96pp; English.  
CC The sequence is that of an antimicrobial protein which can  
be used to control microbial infestations in plants and mammalian  
CC animals.  
SQ Sequence 666 AA;

Query Match 28.8%; Score 1246; DB 1; Length 666;  
Best Local Similarity 39.5%; Pred. No. 3.18e-97;  
Matches 223; Conservative 150; Mismatches 144; Indels 47; Gaps 29;

DB 121 DPQOQYECQRCORRETEPRHMOICQRCERR-EKEK--R-K-Q--QKRYEE-QQ--R 170  
QY 36 DPPKRYEDCRRCMDTRGQEQOQESCKSQYGEKQDQQRHNPEDQRRYEEOQDCR 95  
DB 171 EDEK-----YEER-MKEDNKRDPQREYEDCRHRCQEQEERLQ--HQCQLRCEQOQNH 223  
QY 96 QOERQOPOCQOQRLKRFEOQOQSOFOEQOQHCOEQERPEKQOCVCECKEYQEN 155  
DB 224 GRGDDLMNFORGSGRGYEEGEESQDNFYDERSLSTRFTEGHSVLNFTYGRSKTL 283  
QY 156 PWRGR-EEEAEEET-EEGQEOGSHNPFHRSFQSRFDEHGNFVLRFPASRHL 213  
DB 284 RALKNYRLVLEANPNVFLPHLDADAILLVIGRGALKMIHNDRESYNLECGDVIRI 343  
QY 214 RGINEFRSLIEANPNFVLPHHCDAEKILVLTNGRGLTFLEHKNESYNIIVGVVVKV 273  
DB 344 PAGTFYLINDNNEERLIAKFLDTISPGQYKEFFPAGGNPEPYLSTFEKELLEALN 403  
QY 274 PAGSTVYLANDNKEKLIIVLHRRVNNPGQFEFFPAGSORPSYLAFSTELLEAFN 333  
DB 404 TQTEKLGVCF--Q--QEE-G-VIIR-ASQOIRELTRDSESRHWHIRGESSRGPYN 456  
QY 334 TRSQDLDELFGGRSRRQOQGMFRKASQOIRALSGEATSPRE---KSGE--RFAFN 387  
DB 457 LFNKRLYSNKGAYEVKPEDYRQLODMDSVFIFANTQSGMGPFFNTSTKYVVAS 516  
QY 388 LLSQTPRYSNONGRAFFECAPPEFRQLRDINTVSALQNGSIFVPHYSKATFVILVTE 447  
DB 517 GEADVEMACPIL---SG--RHGGRGGKRRHEED--V-HEOYRARISKEATVYLAGH 568  
QY 448 GNGYAEVSPILPROSSTEEDEEEDQDEERSSGQYRKIRSRISRDIDVVPANF 507  
DB 569 PVTVAASGNENLLFAFGINAQN--N--H-EN-FLAGRENNVLOQIEQAMELATAAPAKE 623  
QY 508 PVTVAASQNLMTGGLYONINPNHONIRIFVAGKINHY-RQWDSQAKELATGVSSRL 566  
DB 624 VEESNSQDSIFFPGRPOHQOOS 647

QY 567 VDEIFNSPQESTFVS-RQORAS 589

## RESULT 5

ID W62830 standard; Protein; 625 AA.  
AC W62830; 27-OCT-1998 (first entry)  
DE Macadamia integrifolia antimicrobial protein.  
KW antimicrobial protein; infestation; control.  
OS Macadamia integrifolia.  
FH key Location/Qualifiers  
FT Peptide 1..28  
FT /note= "signal peptide"  
FT Protein 29..666  
FT /note= "mature protein"  
PN W09827805-A1.  
PD 02-JUL-1998.  
PF 22-DEC-1997; AU0874.  
PR 20-DEC-1996; AU-004275.  
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.  
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;  
DR WPI: 98-37279/32.  
DR N-PSDB: V42316.  
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -  
useful for controlling microbial infestations of plants or mammals  
PS Claim 1; Page 43-45; 96pp; English.  
CC The sequence is that of an antimicrobial protein which can  
be used to control microbial infestations in plants and mammalian  
CC animals.  
SQ Sequence 625 AA;

Query Match 28.7%; Score 1244; DB 1; Length 625;  
Best Local Similarity 39.9%; Pred. No. 4.80e-97;  
Matches 225; Conservative 149; Mismatches 143; Indels 47; Gaps 31;

DB 80 DPQOQYECQRCORRETEPRHMOICQRCERR-EKEK--R-K-Q--QKRYEE-QQ--R 129  
QY 36 DPPKRYEDCRRCMDTRGQEQOQESCKSQYGEKQDQQRHNPEDQRRYEEOQDCR 95  
DB 130 EDEK-----YEER-MKEDNKRDPQREYEDCRHRCQEQEERLQ--YQCQRCQEQOQNH 182  
QY 96 QOERQOPOCQOQRLKRFEOQOQSOFOEQOQHCOEQERPEKQOCVCECKEYQEN 155  
DB 183 GRGDDLMNFORGSGRGYEEGEESQDNFYDERSLSTRFTEGHSVLNFTYGRSKTL 242  
QY 156 PWRGR-EEEAEEET-EEGQEOGSHNPFHRSFQSRFDEHGNFVLRFPASRHL 213  
DB 243 RALKNYRLVLEANPNVFLPHLDADAILLVIGRGALKMIHNDRESYNLECGDVIRI 302  
QY 214 RGINEFRSLIEANPNFVLPHHCDAEKILVLTNGRGLTFLEHKNESYNIIVGVVVKV 273  
DB 303 PAGTFYLINDNNEERLIAKFLDTISPGQYKEFFPAGGNPEPYLSTFEKELLEALN 362  
QY 274 PAGSTVYLANDNKEKLIIVLHRRVNNPGQFEFFPAGSORPSYLAFSTELLEAFN 333  
DB 363 TQTEKLGVLG--Q--QEE-G-VIIR-ASQOIRELTRDSESRHWHIRGESSRGPYN 415  
QY 334 TRSQDLDELFGGRSRRQOQGMFRKASQOIRALSGEATSPRE--E--KSGE--RFAFN 387  
DB 416 LFNKRLYSNKGAYEVKPEDYRQLODMDSVFIFANTQSGMGPFFNTSTKYVVAS 475  
QY 388 LLSQTPRYSNONGRAFFECAPPEFRQLRDINTVSALQNGSIFVPHYSKATFVILVTE 447  
DB 476 GEADVEMACPIL---SG--RHGGRGGKRRHEED--V-HEOYRARISKEATVYLAGH 527  
QY 448 GNGYAEVSPILPROSSTEEDEEEDQDEERSSGQYRKIRSRISRDIDVVPANF 507  
DB 528 PVTVAASGNENLLFAFGINAQN--N--H-EN-FLAGRENNVLOQIEQAMELATAAPAKE 582  
QY 508 PVTVAASQNLMTGGLYONINPNHONIRIFVAGKINHY-RQWDSQAKELATGVSSRL 566  
DB 583 VEELFNSQDSIFFPGRPOHQOOS 606

Query	567	VDEIFNSNPQESFYVS-RQRRAS	589
Db	624	VEELFNSODESIFPPGPRQDOOS	647
Oy	508	PVTIYASQNLRLMTGGLIYONINPDHNOITFVAGKINHV-RQMSQAKELAFGVSSRL	566
Oy	569	PVVFVSSGSENNKLLIFAGSINQON-N-H-EN-FLAGRENNVLOOIEQOAMELFAAPRKE	623
Db	448	GNCGVAEWSPHLPROSSYEEDDEDEDEQOEDEERSQYRKIRSTRLSGDIIVPAPNF	507
Oy	517	GEADVACAPHL---SG-RHGGRGGR-KHEEED--V-HYEOVKAPLSREALIVPVGH	568
Db	388	LLSQTPRRSNOGNRFFECAPREFPOLRDINTVTSALQNGSIVPPIYHNKATFVILYTE	447
Oy	274	PAGSVIVYLANQDNKEKLIIVLHPRVNNPGQFEFFPAGSORPSYLRASRLTELEPAFN	333
Db	404	TOAEKLRGVLG-Q-QRE---GYIIASQOIEQLRDSSESRMRHRRIGESGSRPNY	456
Oy	334	TRSELDLDELFEGRSRRRQOGQMFRRASQOIALISQEAAT-SPR-E--KSGE--RFAFN	387
Db	457	LEFKRPLISNKGAYEKKPEDEYQLOLDMQVSEIANTIGQSNMGPFENFTRSTVVVAS	516
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	224	GRGGLINPQNGSGGRVEEGEKSQNDPIYDENSLSTRPTEGCHLSYLENTYGRSKLL	283
Oy	156	PWRGE-REEEAEEET-EEGQEQSHNPFHFRSFSQFREHEHGNPRVLORFASRPIL	213
Db	171	EDEEK---YEER-MKEEDNRKDQOEYEDRCRRCEQOEPRQO-YOCORRCHEQORQ	223
Oy	96	QOEERQPOCQORCRLKRFPEOEOOSOFQEQCHQOEQRPKQCVRECREKYEQN	155
Db	121	DPOQYEQCQRCQRHREPRHMTCCQRCRRY-EKEK--R-R-Q--QKRYEP-QQ--R	170
Oy	36	DPKPRTEDCRRRCWMDTRGQKEQOQCEESCSQYGEKQDQQRHNEPDRKRECEQOECR	95
Db	284	RALKRYLVLLLEAPNPAFTVPLTHLDADAIIYLTGGRALKIHHHNDRESYNLECGDYIRI	343
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	344	PAGTFYVILNRDNNEKRLIAKFLQITSTPGQYKEFPAGGONPEPVLSTFSKETLEALN	403
Oy	274	PAGSVIVYLANQDNKEKLIIVLHPRVNNPGQFEFFPAGSORPSYLRASRLTELEPAFN	333
Db	404	TOAEKLRGVLG-Q-QRE---GYIIASQOIEQLRDSSESRMRHRRIGESGSRPNY	456
Oy	334	TRSELDLDELFEGRSRRRQOGQMFRRASQOIALISQEAAT-SPR-E--KSGE--RFAFN	387
Db	457	LEFKRPLISNKGAYEKKPEDEYQLOLDMQVSEIANTIGQSNMGPFENFTRSTVVVAS	516
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	224	GRGGLINPQNGSGGRVEEGEKSQNDPIYDENSLSTRPTEGCHLSYLENTYGRSKLL	283
Oy	156	PWRGE-REEEAEEET-EEGQEQSHNPFHFRSFSQFREHEHGNPRVLORFASRPIL	213
Db	171	EDEEK---YEER-MKEEDNRKDQOEYEDRCRRCEQOEPRQO-YOCORRCHEQORQ	223
Oy	96	QOEERQPOCQORCRLKRFPEOEOOSOFQEQCHQOEQRPKQCVRECREKYEQN	155
Db	121	DPOQYEQCQRCQRHREPRHMTCCQRCRRY-EKEK--R-R-Q--QKRYEP-QQ--R	170
Oy	36	DPKPRTEDCRRRCWMDTRGQKEQOQCEESCSQYGEKQDQQRHNEPDRKRECEQOECR	95
Db	284	RALKRYLVLLLEAPNPAFTVPLTHLDADAIIYLTGGRALKIHHHNDRESYNLECGDYIRI	343
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	344	PAGTFYVILNRDNNEKRLIAKFLQITSTPGQYKEFPAGGONPEPVLSTFSKETLEALN	403
Oy	274	PAGSVIVYLANQDNKEKLIIVLHPRVNNPGQFEFFPAGSORPSYLRASRLTELEPAFN	333
Db	404	TOAEKLRGVLG-Q-QRE---GYIIASQOIEQLRDSSESRMRHRRIGESGSRPNY	456
Oy	334	TRSELDLDELFEGRSRRRQOGQMFRRASQOIALISQEAAT-SPR-E--KSGE--RFAFN	387
Db	457	LEFKRPLISNKGAYEKKPEDEYQLOLDMQVSEIANTIGQSNMGPFENFTRSTVVVAS	516
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	224	GRGGLINPQNGSGGRVEEGEKSQNDPIYDENSLSTRPTEGCHLSYLENTYGRSKLL	283
Oy	156	PWRGE-REEEAEEET-EEGQEQSHNPFHFRSFSQFREHEHGNPRVLORFASRPIL	213
Db	171	EDEEK---YEER-MKEEDNRKDQOEYEDRCRRCEQOEPRQO-YOCORRCHEQORQ	223
Oy	96	QOEERQPOCQORCRLKRFPEOEOOSOFQEQCHQOEQRPKQCVRECREKYEQN	155
Db	121	DPOQYEQCQRCQRHREPRHMTCCQRCRRY-EKEK--R-R-Q--QKRYEP-QQ--R	170
Oy	36	DPKPRTEDCRRRCWMDTRGQKEQOQCEESCSQYGEKQDQQRHNEPDRKRECEQOECR	95
Db	284	RALKRYLVLLLEAPNPAFTVPLTHLDADAIIYLTGGRALKIHHHNDRESYNLECGDYIRI	343
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	344	PAGTFYVILNRDNNEKRLIAKFLQITSTPGQYKEFPAGGONPEPVLSTFSKETLEALN	403
Oy	274	PAGSVIVYLANQDNKEKLIIVLHPRVNNPGQFEFFPAGSORPSYLRASRLTELEPAFN	333
Db	404	TOAEKLRGVLG-Q-QRE---GYIIASQOIEQLRDSSESRMRHRRIGESGSRPNY	456
Oy	334	TRSELDLDELFEGRSRRRQOGQMFRRASQOIALISQEAAT-SPR-E--KSGE--RFAFN	387
Db	457	LEFKRPLISNKGAYEKKPEDEYQLOLDMQVSEIANTIGQSNMGPFENFTRSTVVVAS	516
Oy	214	RGINERFRLSILEANPNFTFVLPHHDAEKIYLYVTNGRGLTFLETHENKESINIVPGVYKV	273
Db	224	GRGGLINPQNGSGGRVEEGEKSQNDPIYDENSLSTRPTEGCHLSYLENTYGRSKLL	283
Oy	156	PWRGE-REEEAEEET-EEGQEQSHNPFHFRSFSQFREHEHGNPRVLORFASRPIL	213
Db	171	EDEEK---YEER-MKEEDNRKDQOEYEDRCRRCEQOEPRQO-YOCORRCHEQORQ	223
Oy	96	QOEERQPOCQORCRLKRFPEOEOOSOFQEQCHQOEQRPKQCVRECREKYEQN	155
Db	121	DPOQYEQCQRCQRH	

Query	567	VDEIFNSNPQSEYFVS--RQQRAS	589
RESULT	7		
ID	W62838	standard; Protein; 605 AA.	
AC	W62838		
DT	27-OCT-1998	(first entry)	
DE	Glycine max antimicrobial protein.		
KW	antimicrobial protein; infestation; control.		
OS	Glycine max.		
PN	W09827805-A1.		
PD	02-JUL-1998.		
PF	22-DEC-1997; AU0874.		
PR	20-DEC-1996; AU-004275.		
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.		
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;		
DR	WP1: 98-37279/32.		
FT	Novel anti-microbial protein from e.g. Macadamia integrifolia -		
PT	useful for controlling microbial infestations of plants or mammals		
PS	Claim 1; Page 63-65; 96pp; English.		
CC	The sequence is that of an antimicrobial protein which can		
CC	be used to control microbial infestations in plants and mammalian		
CC	animals.		
SO	Sequence	605 AA;	
Query Match	22.3%;	Score 966;	DB 1; Length 605;
Best Local Similarity	34.2%;	Pred. No. 3,32e-72;	
Matches	133;	Conservative 168;	Mismatches 164; Indels 40; Gaps 28;
Db	50	RN0ACHARNL-LVYKKE--CEGELTPRRPRQHREPERPOQGEKEDEDEOPRRPFP	106
Qy	40	RYE0CRRCCEMDTIGQKREQ0QCESSCSQYGEKDOQRHREPERQRYECC0QCEKQOE-	98
Db	107	PR0PQROEEH0E0R0E0E0WPRKEKREKGESEDEDEDE0E0DER0PPRP0P0K0EEN	166
Qy	99	ER0PQ0C0Q0RLKRF0E0-Q0Q0SR0Q0E-C0Q0CH0Q0E0R0P0E0K0-Q0CYR0E0K0EN	155
Db	167	EEDDEDE0Q0RESESEDESELRRHKKNPFLGSRNRETFLEKNOYGRIVL0R0EN0RSPO	226
Qy	156	PWRGRREEREA-EEEETEEGE-QE0SH-NPEFHRRSF0SRFREHGNFRLV0R0FASRHP1	212
Db	-227	L0NLRDYLLEFNKSPNTLLPNTADADYLVILNGTAISLVNDD0RDSYR0SSGDLR	286
Qy	213	LRGIERFSLILENPNTEFVLPHCDEKILTVNGSGTLFTLHKESINIVPGYVK	272
Db	287	VPSGTTVVVPPDNNENLRITLLIPIVKNQGRSEFSLSSTEA0QSYL0GFSRNILEASY	346
Qy	273	VPAGSTVYLANODKKEKILIAVLRHPNPNPQ0FEFFPAG0RPQSYLRAFSRILEPAF	332
Db	347	DTKREELNKVLFSEDEG0Q0E0LQ0ESYVEIKED0IRALSKAKSSSKRTISSEKPF	406
Qy	333	NTRSE0QDE-LFEGR0SRRR-Q0G-Q-GMPKASQ0E0IRALSOBA-TSPRKK-SGERFAF	386
Db	407	NLRSDPFYSKRLKFEETPEKPNQ0LRDIDFISYDMMEGALLHPNKAIVIVIN	466
Qy	387	NLS0TPRYSN0NRFFEDACPPEPR0LRDINIVYSAL0N0GSIFFVHYNSKAFVILVT	446
Db	467	EGDANIELVG--L-K---EQ0Q0E-Q0E0PLEV-RK---YR--AELS0DIFVIAG	511
Qy	447	EGNGAEVYSPLPROSSYEDEEEDDEE0E0EERRSQYKRIKRSLRGDIFFVYAN	506
Db	512	XPVY-VNATS-NL---NF--FALINMENN0RNFLLAS0DNVISOIV0ELAFPSAQ	564
Qy	507	FPVTFVVS0N0NLMTG0GLY0NIN0INDH0RILFVACKINH-V-R0WDS0AKELAFGVSSR	565
Db	565	AVETLLK-NQRESYFVDAQPKKEE	588
Qy	566	LVDEIFNSNPQSEYFVSQRQRASE	590
RESULT	8		
ID	W62835	standard; Protein; 593 AA.	
AC	W62835		









